ASCO GUIDELINES

Data Supplement

Recommendations for Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology /College of American Pathologists Clinical Practice Guideline Update

Table of Contents

Data Supplement 1: 2013 Update Rationale and Background Information

In 2007, a joint Expert Panel convened by American Society of Clinical Oncology (ASCO) and College of American Pathologists (CAP) met to develop guidelines for when and how to test for the human epidermal growth factor receptor 2 gene (*HER2*, also referred to as *ERBB2*), [1,2] which is amplified and/or overexpressed in approximately 15-20% of primary breast cancers. Since then, minor clarifications and updates to the HER2 Testing Guideline have been issued.[3-5]

This unprecedented collaboration was triggered by the substantial therapeutic benefit observed in most of the first generation, randomized phase-3 clinical trials of trastuzumab in HER2-positive disease, and led to the approval in the adjuvant setting of this humanized monoclonal antibody directed against the extra-cellular domain of the HER2 protein. Trastuzumab had previously been shown to improve progression-free and overall survival (PFS/OS) when combined with chemotherapy in the metastatic setting, even though most patients from the control crossed over to the investigational arm. Since 2005, several of the first generation adjuvant trials have been updated and confirmed the disease-free and overall survival (DFS/OS) benefit offered by one year of trastuzumab administered with or after adjuvant chemotherapy. Recent reported results have suggested that 12 months is the optimal duration of therapy of adjuvant trastuzumab.

Other HER2-targeted drugs (eg, lapatinib, the antibody pertuzumab, and the antibody-drug conjugate ado-trastuzumab emtansine [T-DM1]) have been approved for the treatment of HER2-positive metastatic breast cancer. At the same time, level-1 evidence shows that lapatinib (when added to paclitaxel) and pertuzumab (as a single agent) offer no clinical benefit in patients with HER2-negative metastatic disease. These new HER2-targeted drugs are now being tested in the adjuvant setting, including in studies evaluating their adjuvant role alone or in dual antibody regimens without concomitant or sequential chemotherapy. Therefore, the need for accurate HER2 testing to ensure that the right patient receives the right treatment is now more critical than ever.

In 2007, the ASCO/CAP Panel on HER2 Testing issued guideline recommendations addressing specimen handling and assay performance/reporting aiming to improve the accuracy of HER2 testing for clinical decision-making in breast cancer. The 2007 Panel concluded that available evidence supported using either an immunohistochemistry (IHC) or fluorescence in situ hybridization (FISH) test to select patients with HER2-positive disease for therapy with trastuzumab, and recent data have further supported this recommendation. The 2007 Panel also proposed firm definitions of "HER2-positive" and "HER2-negative" for each assay platform, and established "HER2-equivocal" categories for the two tests. In this regard, the sole purpose of establishing an equivocal category was to trigger reflex testing using the alternative assay to provide clinicians and patients with additional information for clinical decision-making, and not to exclude patients who would otherwise have qualified for participation in the adjuvant trials.

Most importantly, the 2007 ASCO/CAP Panel strongly recommended that laboratories that performed HER2 testing should participate in regular laboratory inspections and bi-annual proficiency testing, such as the testing system established by

CAP. Since the publication of the 2007 HER2 Testing Guideline and the CAP requirement that all its accredited labs use it, there was a notable uptake of proficiency testing (Fig 4), with nearly 1500 laboratories currently participating. CAP also observed fewer laboratories experiencing deficiencies on laboratory inspection. Indirect evidence suggests that the performance of labs that conduct HER2 testing in the US and elsewhere is improving. Intensive educational efforts may have influenced uptake of guideline recommendations, and surveys of participants in CAP educational programs show that over 85% reported changing their HER2 and ER/PgR testing practices. Most importantly, the 2013 Update Committee was pleased with anecdotal evidence suggesting greater communication among all health care providers (including administrators) and the potential positive impact on multidisciplinary efforts to improve the quality of predictive marker testing.

Since 2007, retrospective, hypothesis-generating exploratory data from two of the adjuvant trastuzumab trials suggested that the benefits of adjuvant trastuzumab might have extended to patients whose breast cancers would not have been classified as HER2-positive on repeat central laboratory testing, according to the criteria originally employed for eligibility to the first generation of adjuvant trastuzumab trials (IHC 3+ or FISH ratio ≥ 2.0). However, these tumors were not representative of all HER2-negative tumors as these retrospective observations were made in tumors that had been originally deemed HER2-positive by a local and/or trial reference laboratory at the time of study enrollment. This may, in part, be explained by issues of tissue handling and test interpretation criteria before 2005, or tissue heterogeneity. The NSABP B-47 Trial (NCT01275677) is one study that is prospectively testing this hypothesis in a randomized adjuvant trial of chemotherapy plus or minus trastuzumab in node-positive or high-risk node-negative breast cancer that expresses low levels of the HER2 protein (IHC 1+ or 2+) without amplification.

Along the same lines, a round robin exercise by three expert pathologists in the same central laboratories examined HER2-positive and HER2-negative tissue samples from three first generation trastuzumab clinical trials, and showed a 99% concordance rate for HER2-negative cases. Of interest, HER2 heterogeneity was observed in 5-10% of tumors that tested positive for HER2. At the same time, the overall agreement was only 92% by IHC or FISH at first and, while it improved to 96% upon adjudication, this highlights the need for critical attention to pre-analytical parameters and the potential limitations of existing assay platforms.

The decision by the 2007 Panel to raise the HER2-positive threshold by FISH (*HER2*/CEP17 ratio from 2.0 to 2.2 or HER2 copy number from 4 to 6 copies/cell by IHC (strong circumferential staining from > 10% to > 30% of cells) raised concerns that many patients otherwise eligible for the initial trastuzumab adjuvant trials would have been excluded if using the revised criteria. In fact, a retrospective review of 2,809 patients with HER2-positive disease and available IHC and FISH testing from trial N9831 showed that 0.78% (if first tested by IHC) or 1.1% (if first tested by FISH) of HER2-positive patients would not have qualified for the trial if the ASCO/CAP criteria were used. Using a requirement that circumferential, intense, and complete membrane staining needed to be presented in > 30% cells, and not > 10%, was ultimately estimated to risk incorrectly labeling ~ 0.15% of all newly diagnosed breast cancer as HER2-negative. Nonetheless, after careful deliberation, the 2013 Update Committee agreed it was prudent to revisit the

criterion used to define the HER2-equivocal category, and elected to revert back to the 10% stained cells used for patient entry into the clinical trials.

Since publication of the 2007 Guideline, new diagnostic strategies, like measures of HER2 gene amplification by bright-field in situ hybridization, DNA expression by microarray, or mRNA expression by rtPCR have been introduced into practice, and the Update Committee felt these required evidence-based review. Further experience with established HER2 assays also led to the identification of unusual HER2 genotypic abnormalities, like CEP17 duplication and genomic heterogeneity. Greater insight on the biological and clinical significance of these abnormalities also guided the discussions that were part of this 2013 Guideline Update.

In this regard, the Update Committee has endorsed terminology first suggested by the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative and subsequently adopted by the Institute of Medicine Committee regarding omics-based testing. Semantically, the EGAPP defined three terms: analytical validity, clinical validity, and clinical utility. Analytical validity fundamentally implies that the assay is reliable, accurate, and reproducible, including both pre-analytical and analytical considerations. Clinical validity is shown if the assay separates one population into two different groups that have statistically significant differences in outcomes. Clinical utility, however, is only apparent over time, and if the assay has been shown (with high levels of evidence) to result in improvement in clinical outcomes compared to absence of knowledge of the assay results. Clinical utility may be proposed at the initiation of testing. High levels of evidence can be established either in "prospective, retrospective" studies or in prospectively conducted clinical trials.

The 2013 ASCO/CAP Update Committee wishes to reemphasize that it is important that any new test methodology for the same clinical use be compared with a reference test that assays for the same analyte, and for which there are high levels of clinical utility (i.e., correlations with clinical outcome). For example, if a new test measures DNA amplification for HER2 and has been shown to have high analytical concordance with the predicate test (FISH), the new test is acceptable since amplification of HER2 by FISH has been shown to be associated with benefit from anti-HER2 therapies. A new test that measures an analyte that has not previously been shown to be associated with clinical utility from anti-HER2 therapy by a predicate test must be shown to have both analytical validity and clinical utility according to the definitions used above.

Taken together, these considerations led ASCO and CAP to convene a 2013 Update Committee for a formal and comprehensive review of the peer-reviewed literature published since 2006, and to revise the guideline and recommendations as appropriate. During the deliberations that led to the 2007 Guideline, the Update Committee was equally concerned about inaccurate false positive and false negative HER2 assessments. For example, false positive results would lead to the administration of potentially toxic, costly, and ineffective intravenous adjuvant therapy for a year. The newer agents are also as or more expensive, and may be associated with other dose-limiting toxicities, such as in skin and gastrointestinal tract with lapatinib and in liver with ado-trastuzumab emtansine.

At the same time, the substantial sustained survival benefits of trastuzumab, the long-term low frequency of life-threatening toxicities, and the promise of newer agents/combinations with greater clinical efficacy have led the Update Committee to reexamine the prevalence of false negative HER2 findings. This 2013 Guideline Update

Committee was convened in 2012 to strengthen and clarify the recommendations for HER2 testing in response to novel data and suggestions, and to include newly validated testing platforms such as FDA-cleared assays using bright-field ISH.

Data Supplement 2: Special Issues

2A) Types of Assays for Inclusion

The 2013 Update Committee concluded that there was insufficient evidence to warrant inclusion of mRNA assays (eg, using rtPCR) to determine HER2 status in unselected patients. However, the Update Committee endorses the use of FDA-cleared bright-field ISH assays for the following reasons: (a) the test is measuring a parameter (gene amplification) with demonstrable clinical utility to identify patients likely to benefit from HER2-targeted therapies; (b) a consistent body of evidence shows that bright-field ISH has high concordance levels with other ISH methods using fluorescence (FISH) to measure HER2 gene amplification from ring studies, cohort studies, and external quality assessment (EQA) schemes; and (c) assays appear reproducible across sites. If a CLIA-certified laboratory wishes to use an LDT form of bright-field ISH, the assay must be analytically validated in the laboratory using it and documentation of the clinical validity of the assay must be available.

2B) Polysomy

Polysomy occurs when an entire chromosome is duplicated one or more times, while monosomy is the result of complete deletion of a chromosome. The clinical importance of polysomy or monosomy in the absence of HER2 overexpression on IHC is unknown. However, it is very unlikely monosomy would affect patient care, since if anything it should result in lower HER2 expression. However, polysomy may increase HER2 expression and therefore needs further consideration. Many papers that were identified in the systematic review of the current literature address this issue, but interpretation of results is hampered by different definitions of polysomy. However, evidence thus far examined has shown that there is no relationship of polysomy to HER2 protein status or benefit from HER2-targeted therapy. Polysomy appears to be more common than monosomy, but array-based comparative genomic hybridization (aCGH) studies recently showed that true polysomy (duplication of the entire chromosome) is actually rare. However, loss or gain of the pericentromeric region of chromosome 17 is more commonly observed and can result in alterations in the *HER2*/CEP17 ratio and false-positive or false-negative results. This update has been modified to avoid such false results. (See Table 1 and Data Supplement 2e)

2C) Heterogeneity on ISH

Many new papers have described heterogeneity as a more frequent occurrence than previously thought and identify a subset of patients that might benefit from HER2-targeted therapy. Genomic heterogeneity refers to conditions where more than one population of tumor cells exists within the same tumor. This can occur in three separate manners: as discrete populations (clones) of amplified and non-amplified tumor cells, diffuse intermingling of amplified and non-amplified cells across the tumor, or as isolated

amplified cells in a predominantly non-amplified tumor. Experts contend that the only type of amplified population that is significant is the discrete aggregated cells. Most commonly, these separate populations are randomly counted which could result in inconsistent test results when tested by ISH. Therefore, the 2013 Update Committee recommends a standardized method for ISH interpretation that includes scanning of the entire slide prior to counting and/or using an IHC HER2 test to define areas of potential amplification. Any aggregate population of amplified cells comprising > 10% of the total tumor cell population on the slide must be separately counted. The number of CEP17 and HER2 signals should be counted in a minimum of 20 non-overlapping and contiguous invasive cancer cell nuclei in at least 2 tumor areas of each population of tumor cells (unamplified and amplified areas). The *HER2*/CEP17 ratio should then be calculated for each population of cells individually including the average HER2 signals/cell and ratio of HER2 signals/CEP17 signals, if available. Cases containing amplified *and* non-amplified areas (using this definition) should be reported as positive for HER2. The percentage of the total tumor population with amplification should also be reported.

2D) Consideration for Mandatory Retesting of All HER2-Negative Tests

Several members of the Update Committee considered the option of mandatory retesting for all cases whose initial HER2 test was negative. The Update Committee considered two key elements during its deliberations. The first was the risk of an inaccurate test result by chance alone. For instance, a test that is 95% accurate would see its inaccuracy rate jump from 5% to almost 10% by chance alone upon repeating the same test on the same specimen (0.95 * 0.95 = 0.9025). Second, much of the available data on dual testing came from existing trials in which the second test was performed on blocks collected from patients whose tumors were deemed HER2-positive at the initial institution and who were enrolled in adjuvant trials. In the general population at large, the underlying prevalence of HER2-positive disease is ~ 15-20%. (For more information on the Update Committee members deliberations see Data Supplement 1.) The Update Committee tasked its members to identify existing datasets where all patients in the general population (no pre-selection) had their tumors dual tested. One such example was the Australian In Situ Hybridization Program that reported dual testing data with chromogenic or silver ISH and with IHC in 53,402 cases from 26 laboratories between 2006-2010, and it offered two important pieces of evidence. The first one was a decrease over a 4-year period in the HER2-positivity rate in early stage breast cancer from 23.9% to 15.2% - before laboratory accreditation was increased to 14.6%, after a nationwide system of accreditation was instituted. The other was their observation that the frequency of FISH amplified results among the ~ 31% of all patients with an IHC 1+ (negative) HER2 test decreased from 1.9 % (2009) to 1.1% (2011), or from 0.6% to 0.005% among all patients with dual testing (n= 38,803) during this most recent period with available data.

These data suggest that with proficiency testing and accreditation systems, the risk of false positive testing with either assay declines considerably. Consequently, the Update Committee concludes that the true frequency of a false-negative IHC test in a population-based study may be well below the intrinsic analytic variability of existing HER2 assays in

clinical use, and did not support a recommendation for reflex testing of all patients with an IHC 1+ result as the initial HER2 testing.

Smaller datasets from several investigators appeared to suggest that it might be possible to identify subsets where the level of suspicion of false negativity is markedly raised. However, many of these criteria are consistent with true triple-negative disease, and the Update Committee was unsure whether re-testing was indicated for all such cancers. Also, investigators at the Royal Marsden observed an excellent concordance among 336 patients with early stage breast cancer with paired samples (core biopsy and excision) when tested in both samples for ER, PgR and HER2 (using IHC as initial test) and a discordance frequency of 1.8%, 15%, and 1.2%, respectively. Therefore, the Update Committee was unable to identify a specific subgroup that would benefit from mandatory reflex testing if IHC is less than 2+.

The modified algorithms clarify the previously used "equivocal category" (i.e., HER2 test results that are not clearly positive or negative) and recommends mandatory additional tests to attempt to more definitively determine the HER2 tumor status (positive or negative). The term HER2 "indeterminate" has also been more clearly defined and now applies exclusively to cases where specimen or assay analytical issues prevent the proper determination of the HER2 test result, in which case testing using another specimen should be considered. New types of IHC tests as well as in situ hybridization tests are now approved for use, involving various bright-field illumination methods.

2E) Interpretation Criteria if Using a Dual Signal HER2 Assay and Average HER2 Copy Number < 6 signals/cell (Complement to Figure 3)

1) *HER2*/CEP17 ratio ≥ 2.0 (average *HER2* copy number ≥ 4.0 and < 6.0 signals/cell and average CEP17 = 2.0 signals/cell):

Most specimens with ratios < 3 fall into this signal category (eg, average of five *HER2* signals and average of two CEP 17 signals), and these are the specimens usually referred for a second pathology opinion. These HER2-positive cases are less common, often stain 2+ by IHC, and a small number of subjects in trial N9831 fell in this category. Despite the uncertainty regarding their potential for clinical benefit, these patients were eligible for the first generation of trastuzumab trials.

In regards to clinical outcome, among 1,639 subjects in trial N9831 whose tumors had a *HER2*/CEP17 ratio ≥ 2 by FISH, there were 1,408 IHC 3+ cases with a hazard ratio (HR) of 0.45 (95% CI, 0.32 to 0.63) for improved disease-free survival (DFS) favoring trastuzumab. However, the HR for the remaining 219 cases that had a *HER2*/CEP17 ratio ≥ 2, but were IHC ≤ 2+, was 1.11 (95% CI 0.36-3.43). Despite this retrospective assessment suggesting an apparent lack of a DFS benefit from trastuzumab, the Update Committee agrees with the FDA statement in the US package insert of trastuzumab that "Definitive conclusions cannot be drawn regarding efficacy within other subgroups due to the small number of events" in the first generation of trastuzumab trials.

Consequently, as all patients centrally tested and shown to have a *HER2*/CEP17 ratio ≥ 2 were eligible for the first generation of trastuzumab trials, the Update Committee agrees that a tumor with a *HER2*/CEP17 ratio ≥ 2.0 and an average HER2 copy number ≥

4.0 and < 6.0 signals/cell should be considered HER2-positive without the need for further HER2 testing.

2) *HER2*/CEP17 ratio ≥ 2.0 (average *HER2* copy number < 4.0 signals/cell and average CEP17 < 2.0 signals/cell):

These cases are rare and could represent chromosome 17 monosomy. In such cases, an average CEP17 signal of < 2.0 (or even < 1.5) could result in a *HER2*/CEP17 ratio ≥ 2.0 by ISH, despite a low average HER2 copy number < 4.0. Some Update Committee members have expressed their view that using the HER2:CEP17 ratio alone could be misleading in cases with CEP17 gains or losses, and could lead to an under- or over-estimation of HER2 amplification, respectively. A small number of patients with a *HER2*/CEP17 ratio ≥ 2.0 and an average HER2 copy number < 4.0 signals/cell participated in the first generation trials of adjuvant trastuzumab and, despite their low number, a consideration of them provides some information. An analysis of the impact of HER2 ratio or HER2 copy number on disease outcome by the HERA trialists included 48 patients with tumors showing an average < 4 HER2 copies per cell.[6] These cases necessarily had to have been shown to have HER2 ratio ≥ 2.0 to be eligible for the trial. Given their low copy number, almost all of these tumors would have been among the 453 cases with HER2 ratio ≥ 2.0 and ≤ 4.0. This latter group showed a hazard ratio (HR) for the impact of trastuzumab of 0.56 (95% CI 0.33-0.94) compared with the overall HR of 0.64 (95% CI 0.54-0.76). Thus there was no indication that this group, c.10% of which were cases of < 4 average HER2 copy number, showed any reduced benefit. While an analysis of the 48 cases alone is inappropriate because of its small size, there is no trend in these data to suggest they are non-responsive to trastuzumab. Consequently, the Update Committee agreed that test results for cases with HER2 copy number < 4 and ratio ≥ 2.0 should be primarily reported as HER2-positive (see Figure 3), and this recommendation was made easier in view of the favorable safety profile of trastuzumab. At the same time, several members of the Update Committee expressed concern about describing an invasive breast cancer as HER2-positive on the basis of on a single HER2 test showing a *HER2*/CEP17 ratio ≥ 2.0 and an average HER2 copy number < 4.0 signals/cell and recommended further testing of cases of this type.

This example demonstrates the potential importance of describing the raw ISH data (using cell count, absolute HER2, and CEP17 signals) in the pathology reports to aid in interpretation. This issue is likely to become more important in view of the growing clinical interest in developing HER2-targeted regimens without chemotherapy, such as dual antibodies (eg, trastuzumab and pertuzumab), antibody/small molecule (eg, trastuzumab and lapatinib), and single-agent HER2 conjugates like ado-trastuzumab emtansine.

3) *HER2*/CEP17 ratio < 2.0 (average *HER2* copy number ≥ 4.0 and < 6.0 signals/cell and average CEP17 > 2.0 signals/cell):

While these cases are uncommon, co-amplification of CEP17 region is occasionally observed in some ISH assays and may lead to a *HER2*/CEP17 ratio < 2.0. While this would suggest lack of *HER2* amplification, cases with an average *HER2* copy number ≥ 4.0 and <

6.0 signals/cell should be considered ISH-equivocal. If coamplification of CEP17 is suspected, laboratories may pursue one of two options:

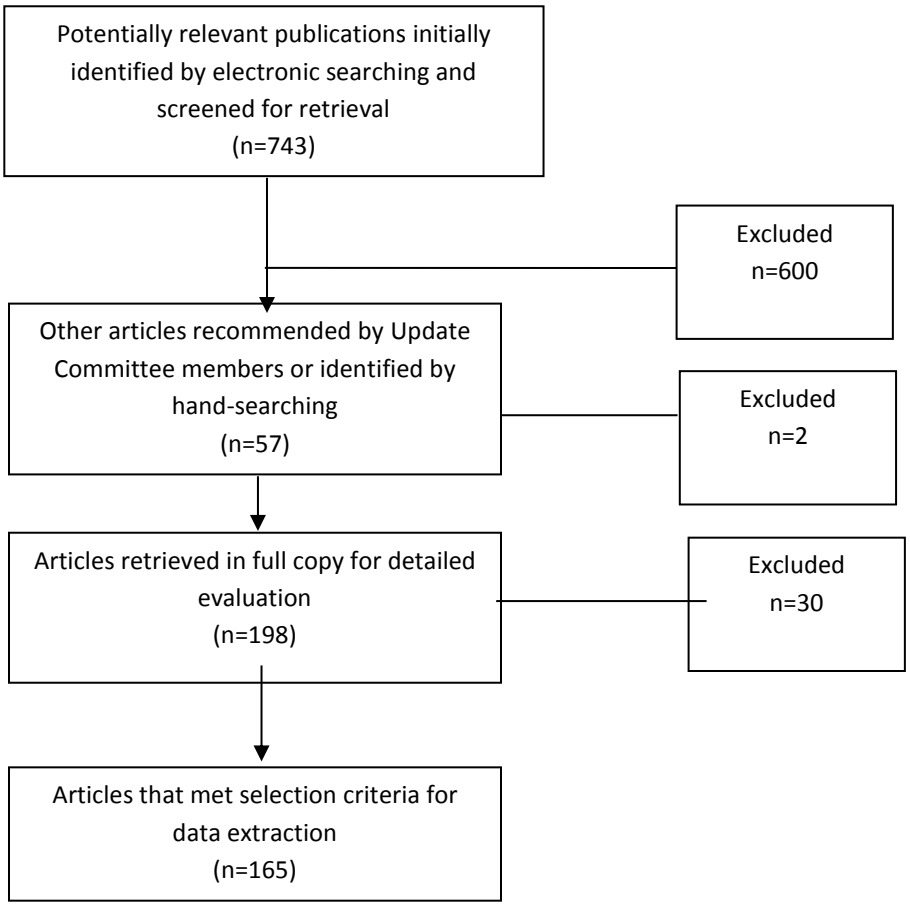
- Repeat HER2 testing in the same specimen using an alternative probe for CEP17 or for another gene in chromosome 17 not expected to coamplify with HER2. If the HER2/CEP17 ratio upon retesting of the same specimen using an alternative ISH probe for chromosome 17 is ≥ 2.0, the tumor should considered HER2-positive. Such cases reinforce the view that high CEP17 copy number is a poor surrogate for chromosome 17 polysomy, and that such cases most likely represent focal pericentromeric gain.[7]
- Perform a reflex HER2 test (IHC or single-signal ISH using same specimen/same block or same specimen/different block) or order a new HER2 test (IHC or ISH, using another available specimen).

Note that in trial N9831, a retrospective assessment, showed 156 subjects whose tumors had *HER2*/CEP17 ratio < 2.0 by FISH with 53 among them showing IHC 3+. While the average HER2 copy number and average CEP17 signals of these N9831 tumors have not been reported, a HR for DFS of 0.57 was observed favoring trastuzumab (95% CI 0.08 to3.89). Of interest, a HR for DFS of 0.51 favoring trastuzumab (95% CI 0.21 to 1.23) was observed in the remaining 103 subjects whose tumors showed a *HER2*/CEP17 ratio < 2 by FISH with an IHC test result ≤ 2+. However, while this retrospective analysis showed that these N9831 specimens with apparent HER2 negative disease benefited from trastuzumab, it is important to note that these patients had tested HER2-positive at the time of trial enrollment in order to fulfill eligibility criteria.

Data Supplement 3: Search terms – January 2013

(("Immunohistochemistry"[MeSH] OR immunohistochemistry[tiab] OR immunocytochemistry[tiab] OR “IHC”[tiab] OR "In Situ Hybridization, Fluorescence"[MeSH] OR “fluorescence in situ hybridization”[tiab] OR “fluorescence in-situ hybridization”[tiab] OR “FISH”[tiab] OR (chromogenic[tiab] AND hybridization[tiab]) OR “CISH”[tiab] OR ((gold-facilitated[tiab] OR autometallographic[tiab] OR “bright field”[tiab] OR bright-field[tiab]) AND hybridization[tiab]) OR “GOLDFISH”[tiab]) AND (Genes, erbB-2[MeSH] OR Receptor, erbB-2[MeSH] OR “Her-2”[tiab] OR “Her2”[tiab] OR “HER2”[tiab] OR “HER-2”[tiab] OR “erbB-2”[tiab] OR “erbB2”[tiab] OR “epidermal growth factor receptor-2”[tiab] OR “epidermal growth factor receptor 2”[tiab] OR receptor, epidermal growth factor[mh] OR epidermal growth factor receptor-neu receptor[nm]) AND (Breast neoplasms[MeSH] OR “breast neoplasm*”[tiab] OR “breast cancer*”[tiab] OR “breast tumor*”[tiab] OR “breast tumour*”[tiab])) NOT (animals [mh] NOT human [mh])

Data Supplement 4: QUORUM Diagram



Data Supplement 5: Clinical Questions

This guideline update addresses two principal questions regarding HER2 testing.

1. What is the optimal testing algorithm for the assessment of HER2 status?

2. What strategies can help ensure optimal performance, interpretation, and reporting of established assays?
   - Testing analytic validation requirements
   - Ongoing competency assessment
   - Reporting requirements
   - Regulatory framework
   - Optimal external quality assurance methods to ensure accuracy in HER2 testing and laboratory accreditation

Data Supplement 6: Preanalytic Issues

- **Time to fixation (cold ischemic time)**: Because of the potential importance of biomarkers for determining the most appropriate treatment options for certain patients, there is a need for standardizing pre-analytic variables, with the goal of developing standardized methods of tissue procurement and processing, and documenting how these variables affect the quality of tissue for biomarker testing and molecular analysis. Recent reports have suggested that excessive delay from tissue collection to the initiation of formalin fixation has the potential to adversely impact the analysis of hormone receptor assays and HER2 analysis. [8,9] Khoury et al. analyzed 10 resected breast cancers and suggested that delays to the start of fixation of 60 to 120 minutes may compromise the accurate analysis of ER, PgR and HER2 FISH due to loss of staining or hybridization signal intensity. The implications of these findings are that some tumors with excessive cold ischemic times may be falsely classified as negative for the expression of these important therapeutic targets.
- **Duration of tissue fixation**: A number of recent studies have addressed the issue of prolonged tissue fixation, including two prospective validation studies that compared the results of ER, PR and HER2 studies on tissue fixed for a standard amount of time with tissue from the same samples that underwent prolonged fixation (72 to 96 hours). [10,11] The data from these and other studies suggest that formalin fixation for up to 72 hours does not appear to have any impact on ER, PgR and HER2 reactivity and therefore is an acceptable upper limit of time in routine clinical practice. The immune-reactivity of breast prognostic markers testing for ER, PgR and HER2 may be reduced by very long, extended formalin over-fixation that is not clinically relevant. It is important to measure time to fixative (including sectioning) and time in fixative. It is recommended that the time from removal from the patient to incision of the specimen be as short as possible, ideally within 1 hour.
- **Specimen selection for HER2 analysis:** A number of studies comparing ER, PgR and HER2 analysis on core needle biopsies and the subsequent excision sample from the same patient have shown comparable results. [12] Furthermore, some of these reports have suggested that the core needle sample may be a better specimen for analysis because these tissues are usually placed directly into fixative within minutes and the formalin will infiltrate more quickly, resulting in more uniform and consistent tissue fixation. It is important that both specimen types conform to the pre-analytic requirements for the test (particularly adequate fixation time) if they are to be used for HER2 analysis. While the excisional specimen is a more representative sampling of the patient's tumor, the needle core biopsy can be used for HER2 analysis with the following caveats. Repeat HER2 testing on the excisional sample is recommended when 1) the result is negative and the sample is limited on the core biopsy or 2) the results do not fall in the clearly positive or negative range (IHC or FISH) on the core. In such a situation, the test must be repeated on the same or another sample using the alternative method. After this method is completed, the HER2 interpretation can be rendered. See Table 2 for definition of specific situations.

Data Supplement 7: IHC Interpretation Criteria

Review controls; if not as expected, test should be repeated.
More than 10% of tumor must show circumferential membrane staining for positive result.
Membrane staining must be intense and uniform and resemble chicken wire.
Ignore incomplete or pale membrane staining.
Quantitative image analysis is encouraged for cases with weak membrane staining (1-2+) to improve consistency of interpretation among pathologists. Variation in visual acuity, light sources, and microscopes cannot be controlled in manual counting situations.
If cytoplasmic staining obscures membrane staining, repeat assay or do alternative ISH testing
Reject sample if normal ducts and lobules show obvious cytoplasmic staining unless this is in areas of apocrine metaplasia.
Reject sample if there are obscuring artifacts, such as crush or edge artifact.
Avoid scoring DCIS; score only infiltrating ductal carcinoma.

Abbreviations: ISH, in situ hybridization; FDA, US Food and Drug Administration; HER2, human epidermal growth factor receptor 2

Data Supplement 8: ISH Interpretation Criteria
Review corresponding hematoxylin and eosin and/or IHC slide to localize the invasive cancer; carcinoma in situ should not be scored.
Review controls; if not as expected, test should be repeated.
Review entire slide subjected to ISH testing to define whether there is more than one population of cells with variable numbers of signals/cell. This review must be done by a pathologist prior to counting of the sample. Alternatively, perform IHC and use areas of 3+ staining for ISH. For bright-field ISH, compare signals over normal breast to those over tumor cells to assure that normal and amplified areas can be readily distinguished. If there pattern is not clearly either unamplified or amplified on this review, slides should be submitted for expert opinion. With fluorescent and bright-field ISH assay, count at least 20 non-overlapping cells in two separate areas of invasive cancer (at least 10/area). If two populations of tumor cells show different levels of amplification, each should be separately counted including 20 non overlapping cells in each population. The percentage of the entire tumor on the slide that shows HER2 gene amplification must be defined by IHC, by estimation, or by image analysis, and must be reported. Only amplified populations of 10% or more of the entire cell population on the tissue sample should be reported.
Reject if signals are show variable staining intensity (> 25%).
Reject if autofluorescence high or nuclear resolution poor (FISH).
Reject if background obscures signal resolution (> 10% over cytoplasm).
If *HER2*/CEP17 ratio between 1.8 and 2.2, have additional person count an additional 20 non overlapping cells. If *HER2 signals per cell are > 6, assay should be considered positive regardless of ratio.*
Excessive numbers of CEP 17 signals do not influence interpretation but the average number of CEP17 signals per cell should also be provided along with the average number of *HER2* signals/cell.
Counting can be done by a trained technologist, but pathologist must confirm that result (count) is correct and that invasive tumor was counted. Pathologist must survey entire tumor before counting to define whether more than one population of cells is present and the percentage of the tumor that this population represents. This survey may also be done using IHC protein expression to select the area for ISH counting.

Abbreviations: ISH, in situ hybridization; HER2, human epidermal growth factor receptor 2

Data Supplement 9: Reporting Elements for IHC
Patient identification information
Physician identification
Date of service
Specimen identification (case and block number)
Specimen site and type
Specimen fixative type must be recorded on each specimen, but not in report
Time to fixation must be recorded on each specimen, but not in report
Duration of fixation must be recorded on each patient, but not in report
Antibody clone/vendor
Method used (test/vendor and if FDA approved)
Image analysis method (if used)
Controls (high protein expression, low-level protein expression, negative protein expression, internal)
Adequacy of sample for evaluation
Results
Percentage of invasive tumor cells exhibiting complete membrane staining
Uniformity of staining: present/absent
Homogeneous, dark circumferential pattern: present/absent
Interpretation should conform to new guideline recommendations:
HER2 test result positive, HER2 test result negative, HER2 test result equivocal, HER2 test result indeterminate
Comment
Statement should be made in each report about whether specimen handling falls within guideline recommendations. Time to fixation, type of fixative, and fixation duration does not need to be included in the report, but must be documented in the lab accession slip or a similar document. If specimen handling falls outside recommended guidelines, this should be clearly reported in a comment in the pathology report. If an FDA-approved method is used, it should be stated; if the FDA-approved method has been modified, a statement in the report should be included indicating what modifications were made.; if the test is not FDA approved or an FDA-approved test has been used, include a statement that the laboratory follows requirements of CAP or other accreditor for laboratory developed test (LDT) reporting. If the tumor will be submitted for additional testing or if another sample will be testing in addition, this should be included in the comment section. If a decision is made to pursue further testing, this should be documented in the comment section.

Abbreviations: ISH in situ hybridization; FDA, US Food and Drug Administration; HER2, human epidermal growth factor receptor 2

Data Supplement 10: Reporting Elements for ISH
Patient identification information
Physician identification
Date of service
Specimen identification (case and block number)
Specimen site and type
Specimen fixative type must be recorded on each specimen, but not in report
Time to fixation must be recorded on each specimen, but not in report
Duration of fixation must be recorded on each specimen, but not in report
Probe(s) identification
Method used (specifics of test/vendor and if FDA approved)
Image analysis method – manual or automated
Controls (amplified, equivocal, and non-amplified, internal)
Adequacy of sample for evaluation (adequate number of invasive tumor cells present)
Results for each discrete population of tumor cells (defined by having discrete different gene amplification)
Number of invasive tumor cells counted
Number of observers
Average number of HER2 signals/nucleus or tile
Average number of CEP 17 chromosome probes/nucleus or tile if dual probes used
Ratio of average HER2 signals/CEP 17 probe signals if CEP17 probe used
Note: Tile is unit used for image system counting
Interpretation
Interpret HER2 test result positive, HER2 test result negative, HER2 test result equivocal, or HER2 test result indeterminate. If two populations of tumor cells are identified (one amplified and one unamplified), each should be separately reported as positive or negative and the actual counts recorded separately. The amount of the invasive tumor of the amplified population should also be provided. Only amplified populations of 10% or more of the entire cell population on the tissue sample should be reported.
Comment
Statement should be made in each report about whether specimen handling falls within guideline recommendations. Time to fixation, type of fixative, and fixation duration do not need to be included in the report, but must be documented in the lab accession slip or a similar document. If specimen handling falls outside recommended guidelines, this should be clearly reported in a comment in the pathology report. If an FDA-approved method is used, it should be stated; if the FDA-approved method has been modified, a statement in the report should be included indicating what modifications

were made.; if the test is not FDA approved or an FDA-approved test has been used, include a statement that the laboratory follows requirements of CAP or other accreditor for laboratory developed test (LDT) reporting.

If the tumor will be submitted for additional testing or if another sample will be testing in addition, this should be included in the comment section. If a decision is made to pursue further testing, this should be documented in the comment section.

Abbreviations: ISH, in situ hybridization; FDA, US Food and Drug Administration; HER2, human epidermal growth factor receptor 2.

Data Supplement 11: Concordance

Statistical Criteria for Validation

Effectiveness of the HER2 testing guidelines for ensuring reliable HER2 determinations for clinical decision making depends critically on the validity of the assays performed in individual laboratories. Guidance on interpretation of a particular score from an immunohistochemical or in situ hybridization assay relies on the comparability of those scores across laboratories. Each laboratory that performs HER2 testing must validate its HER2 assays to ensure that this comparability requirement is met.

A laboratory performing HER2 assays must operate in an environment conducive to consistent and reliable testing. Requirements include adequate staff with appropriate training, careful documentation of assay protocols and adherence to those protocols, quality monitoring programs, and periodic proficiency testing (for example, participation in the CAP HER2 Proficiency Testing Program). Any assays that have been modified or newly introduced to a laboratory must undergo validation to ensure that they produce results consistent with previously validated assays.

Whether the purpose of the assay validation is to compare an immunohistochemical assay to an in situ assay, a modified immunohistochemical assay to a previously validated immunohistochemical assay, or other such comparisons, validation requires that an adequate number of cases representative of the clinical setting be assessed by both the new and previously validated assays to ensure that there is high concordance, typically 95%, between the two sets of results. An assay that does not achieve acceptably high concordance with established validated assays must undergo a rigorous clinical validation. A new clinical validation would have to be performed on a set of clinical samples for which the necessary pathological and clinical variables, including clinical outcome, are known, and it must be established that the new assay results reliably predict the clinical outcome of interest (eg, clinical benefit from anti-HER2 therapy). A thorough treatment of requirements for a full clinical validation is beyond the scope of this discussion.

First principles dictate that the cases included in the validation set be representative of the types of cases that would be seen in routine practice. The cases should be representative with regard to specimen acquisition and handling, and the cases in the validation set should include a representative range of strong and weak positive and negative results. The inclusion of only clear cut positive and negative cases in the validation set will yield an optimistically biased estimate of concordance.

Guidance on the number of cases that should be in the assay validation set is provided by statistical calculations based on the binomial distribution. The validation criteria could be applied to a mixture of positive and negative cases representative of what would be encountered in clinical practice, or preferably separate validations should be carried out for positive cases and for negative cases (as defined by the previously validated assay). For combined positive and negative cases, one would refer to the *overall concordance rate*; otherwise, the terms *positive concordance* and *negative concordance* refer to agreement on the subset of cases called by the validated assay to be positive and negative, respectively. For a particular concordance rate of interest (overall, positive, or negative), the probability that the observed concordance will meet or exceed a specified

benchmark will depend on the true underlying concordance rate and the number of cases tested. Table 1 displays these relationships.

If 80 cases are tested by both the validated and new assays and a benchmark of 90% concordance is set for declaring successful validation, then one refers to the column corresponding to N=80 in Table 1. The probability is 98% that the observed concordance rate will meet or exceed the 90% benchmark if the true concordance rate is 95%, but only 59% if the true concordance rate is 90%, or 13% if the true concordance rate is 85%. In order to more sharply distinguish between a true concordance of 90% versus 95%, the number of cases in the validation set would have to be increased beyond 80. If the benchmark for validation is set at 95% concordance, then with N=80 the probability is 98% that the observed concordance rate will meet or exceed the 95% benchmark if the true concordance rate is 98%, but only 63% if the true concordance rate is 95%, or 9% if the true concordance rate is 90%. Due to variability in the estimated concordance on a finite number of cases, the observed concordance in a particular testing session may drop slightly below 95%; however, a laboratory should strive for a long term average concordance of 95% or better.

Caution should be used in the interpretation of the overall concordance rate because it is influenced by the proportions of positive and negative cases in the validation sets if the underlying concordance rates comparing validated to new assay differ for positive and negative cases. If half of the 80 cases are positive and half are negative by the validated test, then one can again use Table 1, but refer to the column corresponding to N=40, to set up a validation strategy for positive and negative cases separately. Best practice is to report validation results separately for positive and negative cases, and preferably separately within scoring categories for semi-quantitative tests (eg, 0, 1+, 2+, 3+ for immunohistochemical assays), even if the concordance benchmark is not applied separately to each scoring category.

Table 1: Probability (%) of Meeting or Exceeding Benchmark Concordance Rates of 80%, 85%, 90% or 95% (revised)

These probabilities include the results where the percentage of correct cases *equals* the benchmark rate ("meeting" the benchmark concordance rate).

		Number of Cases in Test Set		
Benchmark Concordance Rate (%)	True Concordance Rate (%)	20	40	80
80	80	63	59	57
	85	83	86	92
	90	96	98	>99
	95	>99	>99	>99
	98	>99	>99	>99
85	80	41	29	16
	85	65	61	58
	90	87	90	95
	95	98	>99	>99
	98	>99	>99	>99
90	80	21	8	1
	85	40	26	13
	90	68	63	59
	95	92	95	98
	98	99	>99	>99
95	80	7	1	0
	85	18	5	0
	90	39	22	9
	95	74	68	63
	98	94	95	98

Data Supplement 12: Examples of International Quality Assurance Program Links

UK NEQUAS http://www.ukneqas.org.uk

CAP http:www.cap.org

RCPA - http://pathologyaustralia.com.au/

ESP European society of pathology - http://esp-pathology.org/ESP

Nordiqc - http://www.nordiqc.org/

REFERENCES – in published manuscript

Additional:

1. **Wolff AC, Hammond ME, Schwartz JN, et al: American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. J Clin Oncol 25:118-45, 2007**
2. **Wolff AC, Hammond ME, Schwartz JN, et al: American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. Arch Pathol Lab Med 131:18-43, 2007**
3. **Hammond EH, Wolff AC, Hayes DF, et al: Reply to G. Sauter et al. J Clin Oncol 27:e153-4; author reply e155-7, 2009**
4. **Hammond ME, Hayes DF, Wolff AC: Clinical Notice for American Society of Clinical Oncology-College of American Pathologists guideline recommendations on ER/PgR and HER2 testing in breast cancer. J Clin Oncol 29:e458, 2011**
5. **Wolff AC, Hammond ME, Hayes DF: Re: predictability of adjuvant trastuzumab benefit in N9831 patients using the ASCO/CAP HER2-positivity criteria. J Natl Cancer Inst 104:957-8, 2012**
6. **Dowsett M, Procter M, McCaskill-Stevens W, et al: Disease-free survival according to degree of HER2 amplification for patients treated with adjuvant chemotherapy with or without 1 year of trastuzumab: the HERA Trial. J Clin Oncol 27:2962-9, 2009**
7. **Hanna WM, Ruschoff J, Bilous M, et al: HER2 in situ hybridization in breast cancer: clinical implications of polysomy 17 and genetic heterogeneity. Mod Pathol, 2013**
8. **Nkoy FL, Hammond ME, Rees W, et al: Variable specimen handling affects hormone receptor test results in women with breast cancer: a large multihospital retrospective study. Arch Pathol Lab Med 134:606-12, 2010**
9. **Khoury T, Sait S, Hwang H, et al: Delay to formalin fixation effect on breast biomarkers. Mod Pathol 22:1457-67, 2009**
10. **Tong LC, Nelson N, Tsourigiannis J, et al: The effect of prolonged fixation on the immunohistochemical evaluation of estrogen receptor, progesterone receptor, and HER2 expression in invasive breast cancer: a prospective study. Am J Surg Pathol 35:545-52, 2011**
11. **al. Y-AIe: in press. American Journal of Clinical Pathology, 2013**
12. **Arnedos M, Nerurkar A, Osin P, et al: Discordance between core needle biopsy (CNB) and excisional biopsy (EB) for estrogen receptor (ER), progesterone receptor (PgR) and HER2 status in early breast cancer (EBC). Ann Oncol 20:1948-52, 2009**